Final Report on "Using Initial Tendency Errors to Reduce Systematic Errors, Identify Model Errors, and Construct Stochastic Parameterizations"

Investigator:	Timothy DelSole
Recipient Name:	George Mason University
Award Number:	NA06OAR4310001
Reporting Period:	1 February 2006- 31 January 2010
Award Period:	1 February 2006- 31 January 2010 (with no-cost extension)
Program Office:	OAR Climate Program Office (CPO)
NOAA Person of Contact:	Hua-Lu Pan, EMC/NCEP/NWS

This report documents our accomplishments during the period 1 February 2006 - 31 January 2010 in pursuit of the research goals stated in the proposal titled "Using Initial Tendency Errors to Reduce Systematic Errors, Identify Model Errors, and Construct Stochastic Parameterizations."

1 Overall Goals and Methodology

The goal of this project was to improve the forecast skill of the GFS and CFS by developing an empirical correction algorithm that would subtract the systematic tendency error at every time step, and by developing stochastic models that would perturb the model in such a way as to produce a forecast ensemble that accounted for both uncertainty in initial condition and model error. The methodology for doing this was two fold. First, the tendency errors would be estimated by fitting the forecast errors at 6, 12, 18, and 24 hour lead times to a function of lead time, and then using the slope of the resulting fit to identify the tendency error growth rate. Then, the climatological mean tendency error would be subtracted from the model integration at every time step. It is important to recognize the innovation in this methodology: by using short forecasts, errors in model components have little time to interact with errors in other components. Second, the residuals from the fit would be used to estimate the statistics of stochastic forcing terms that could be added to the model equations.

2 Summary of Results for the GFS

Tendency errors were estimated for the momentum, temperature, and moisture update equations in the GFS. Importantly, the estimation was performed individually and independently to spectral coefficients, as opposed to grid points, because large scale basis functions are expected to be less noisy and vary more slowly than individual grid points, thereby allowing more accurate estimates of tendency errors. We found that the tendency errors differed across the June 2005 boundary, owing to model changes, so that data only after June 2005 was used to correct the current version of the GFS. Consequently, the tendency errors were

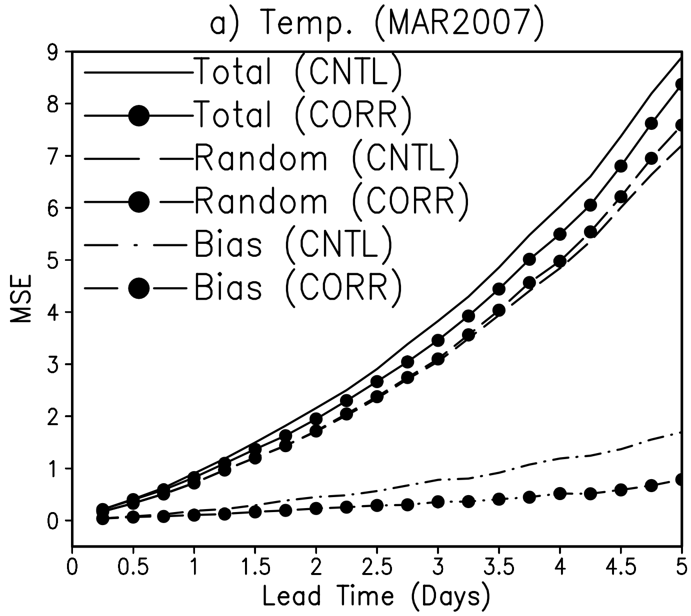

Figure 1: Mean square error of temperature at sigma level 0.2 for March 2007. The error is decomposed into total (solid), bias (dot dashed), and random (dashed) components for control (without mark) and corrected (marked with closed circles) runs.

estimated only during June 2005-June 2007. The tendency error growth rate was fitted to a sinusoid to capture the annual cycle, and this function was then subtracted from the tendency equations to correct the momentum, temperature, and moisture at every time step.

A set of 5-day forecasts, initialized at 0Z at each day in March 2007, were made with both the original GFS and empirically corrected GFS. The mean square error of temperature, averaged over the northern hemisphere on the 0.2 sigma surface, is shown in figure 1 for the control run and the corrected run. The figure shows that the empirically corrected model has less error than the control model. The figure also shows the mean square error decomposed into bias and random components. These latter results reveal that all of the improvement in the mean square error is due to reduction in forecast bias. Other prognostic variables (i.e., zonal velocity, meridional velocity, and moisture) had relatively little bias and thus were only marginally improved, if at all. The same results were obtained for different verification years. The major conclusion to draw from these results is: *empirical correction of the GFS does not improve the forecast above the improvement one would expect from a simple after-the-fact correction, in which the bias of the forecast is subtracted from the forecast.* Since an after-the-fact correction is much easier to perform than empirical correction, we conclude that an empirical correction is not worthwhile for the GFS, at least for improving the forecast. These and other results have been summarized in our paper Yang et al. (2008).

In a separate study (funded by a different federal agency), the principal investigator applied a similar empirical correction method the coupled land-atmosphere model at COLA. The results of the two studies are consistent and can be summarized very simply: a state-independent empirical correction can improve the forecast bias, but leads to no detectable improvement in the random error (i.e, in the non-constant forecast error).

The above conclusion was not anticipated. Therefore, we undertook a systematic study of the literature to understand why empirical correction seems to improve forecast skill in some models but not in others. This study lead to the following hypothesis: *a bias correction can improve forecast skill only if the bias is sufficiently large*. This hypothesis is physically plausible: a large bias implies strong climate drift and relatively useless forecasts, whereas a small bias probably does not degrade the forecast skill. This hypothesis also seems to explain the conflicting results in the literature. On the one hand, Saha (1992) and DelSole and Hou (1999) found that nudging based on tendency errors did not substantially improve skill, but the bias in their models was always less than 10% for time scales less than 5 days. On the other hand, Johansson and Saha (1989), Yang and Anderson (2000), and Danforth et al. (2007) found that empirical correction improved skill, but each of these models appear to have a large bias (we say "appear" because the bias was not reported in these papers). Specifically, Yang and Anderson (2000) found that state-independent nudging improved skill, but their original model could not beat the skill of a persistence forecast for the Niño-3 region in the first 3 months. Similarly, Danforth et al. (2007) found that state-independent correction improved skill, but the bias in their models presumably was large since they used idealized models (such as a quasigeostrophic model) to forecast the NCEP-NCAR Reanalysis. Johansson and Saha (1989) also found that a state-dependent correction improved forecast skill, but their bias fluctuated from being dominant on short time scales to constituting about 30% of the total error after 20 days.

Although the results did not turn out as anticipated, and indicate that the empirical correction method is not a recommended approach for operational models, we believe that the present research serves a very useful purpose. Specifically, we believe that these results represent the most rigorous and careful analysis of empirical correction to date. For instance, the tendency errors were estimated from analyses that were consistent with the dynamical model being corrected, the tendency errors were estimated for individual spectral coefficients, which have much less noise than individual grid points, a significance test was performed to ensure that the correction terms were statistically significant, etc. Thus, the present research compellingly demonstrates that state-independent corrections are unlikely to improve the random error or anomaly skill of operational forecast models with small biases. Such definitive results are useful even if they do not lead to improvements in the GFS and CFS– for instance, they are helpful in assessing future proposals of empirical correction.

3 Results for the CFS

Here we report results of applying the AGCM-correction to the CFS. These results have not been published.

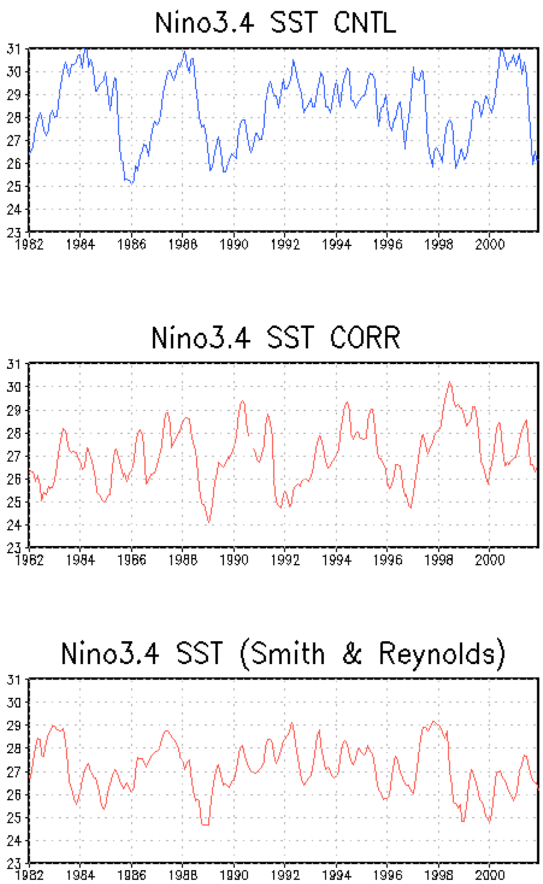

Figure 2: Time series of the NINO3.4 index produced by the control CFS, corrected CFS, and as estimated by Smith and Reynolds (2004). The corrected CFS refers to the CFS model using the empirical correction terms for atmospheric temperature and winds, as derived for the GFS.

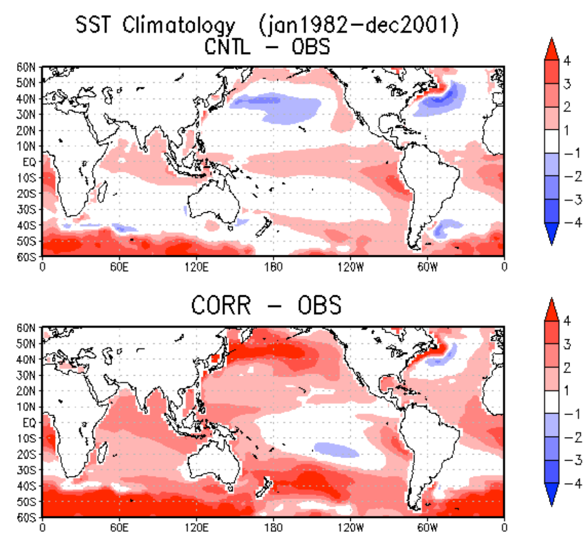

Figure 3: The time mean SST error of the control CFS (top) and the AGCM corrected CFS models, over the period 1982-2001. The corrected CFS refers to the CFS model using the empirical correction terms for atmospheric temperature and winds, as derived for the GFS.

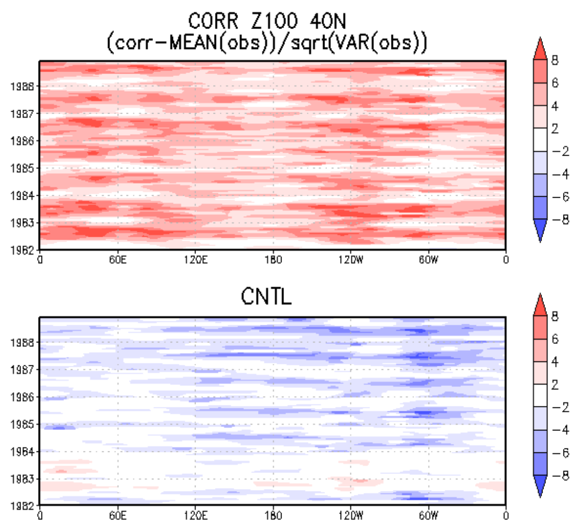

Figure 4: The monthly mean standardized 100hPa geopotential height at 40°N in the AGCM-corrected CFS (top panel) and in the control CFS (bottom panel). The standardized value is the difference between the forecast and observed values, divided by the observed standard deviation.

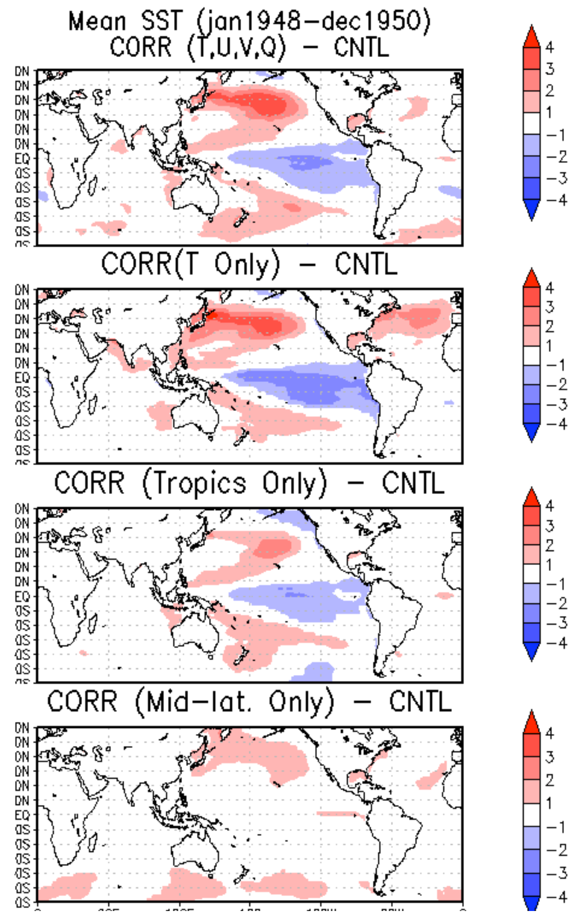

Figure 5: The 3-year mean difference between the corrected CFS and the control CFS for four different variations of the empirical correction: correction of temperature (T), momentum (U, V), and moisture (Q) (top panel), correction of T only (second panel), correction of U, V, T, Q in tropics only (third panel), and correction of U, V, T, Q in midlatitudes only (last panel).

We applied the empirical corrections derived for the GFS to the *atmospheric* component of the CFS, and then ran the model for 20 years. The resulting NINO3.4 index for the control and corrected CFS are shown in fig. 2; for reference, we also show the observed NINO3.4 index in the bottom panel. The figure shows that the correction reduces the NINO3.4 variance and reduces the NINO3.4 bias error, which brings the variability of NINO3.4 more in line with observations. The spatial variation of the bias in SST is shown in fig. 3. The figure shows that the AGCM correction causes a significant warm bias in most ocean locations. The space-time variation of this warm bias in the atmosphere is illustrated in fig. 4, which shows the standardized geopotential height, defined as

$$Z_{standardized} = \frac{Z_{model} - mean[Z_{observed}]}{\sqrt{var[Z_{observed}]}}. \tag{1}$$

The idea for using this variable is that if the variable is normally distributed, then the absolute value of the standardized variable should exceed 1.96 only 5% of the time. We note that the mean and variance depends on month. This standardization allows us to check the consistency of the variable with respect to observations. We see from fig. 4 that the control CFS has a slight cold bias during summer, whereas the empirically corrected CFS has a strong warm bias during the summer. Apparently, the empirical correction overcompensates for the cold bias in the control.

To explore possible reasons for this warm bias, we re-ran the CFS under different empirical corrections. The result of empirically correcting all prognostic variables, temperature only, prognostic variables in the tropics only, and in midlatitudes only, are shown in fig. 5. Comparison of the results shows that the CFS response is dominated primarily by the temperature corrections– that is, the wind and moisture corrections seem to have little impact on the CFS– and most of the temperature correction is achieved by forcing the tropics. We also explored various corrections of surface winds, on the theory that the ocean bias represented a response to wind stress, but found very little impact due to the wind corrections. We discussed these results extensively with COLA scientists and NCEP scientists, but were unable to find ways to avoid the warm bias induced by the empirical correction.

It should be noted that the warm bias discussed above occurs primarily after the first year. In particular, the cold bias in the control during the the first three months is substantially removed by the correction. Thus, there is the possibility that during the first three months the overall forecast skill of the CFS is improved by the empirical correction. To test this possibility, we ran 5-member ensemble forecasts starting in December and ending in March during the 10-year period 1983-1992. The anomaly correlation skills of the control and corrected CFS revealed no statistically significant differences.

It is fair to say that our attempt to improve the CFS using empirical correction methods was "unsuccessful." The significant ocean drift induced by the empirical correction suggests that atmosphere-only corrections to coupled models is not appropriate, because such corrections neglect tendency errors due to the ocean model. It is plausible that the coupled CFS has different tendency errors than the GFS, and that an empirical correction of the full CFS would have yielded better performance in the CFS. However, this approach is much more computationally demanding and was beyond the scope of the proposed research.

4 Rigorous Estimation of Empirical Correction and Stochastic Forcing Terms

Another component of the proposed research was to estimate the empirical correction terms and the stochastic forcing terms within a mathematically rigorous data assimilation framework. This component of the research was especially fruitful, we believe, at least from a basic research point of view. We are especially proud of developing a new filter, called the Diffuse Ensemble Square Root Filter (DESRF), for avoiding "filter collapse" in a mathematically rigorous way. Yang and DelSole (2009a) show that the diffuse filter is superior to the traditional filter in the regime of copious observations, small ensemble size, and imperfect model– a regime that is relevant to atmospheric and oceanic data assimilation. However, the DESRF has a significant practical drawback in that it presently requires inversion of very large matrices, though there exists an equivalent optimization framework that avoids this inversion (which we did not develop further). We also have developed an approach to estimating *multiplicative* model parameters in the ensemble Kalman Filter. This problem is more difficult than estimating additive model parameters because multiplicative parameters change the dynamical structure of the model and may inadvertently cause the model to be unstable, after which the filter blows up. The new methodology is summarized in Yang and DelSole (2009b).

In addition to empirically correcting the GFS and CFS, we had proposed to include stochastic parameterizations in the models. Unfortunately, the available methods for specifying this stochastic parameterization are quite ad hoc (Buizza et al., 1999; Shutts, 2005; Berner et al., 2008, 2009). We saw little value in inserting stochastic terms in the GFS and CFS without a rational method for specifying these terms. Accordingly, we investigated methods for estimating the stochastic model parameters within a rigorous mathematical framework. We were quickly surprised to discover that conventional parameter estimation methods fail to give good estimates of stochastic parameters. This was all the more surprising given that some papers claim otherwise! We developed a Bayesian method for estimating stochastic terms in a stochastic-dynamical model. The new method is based on generalized maximum likelihood estimation, in which the "optimum" estimate is the value that maximizes the distribution of the parameters conditioned on the observations, underlying stochastic-dynamical model, and prior information. The method is equivalent to minimizing a cost function that is similar to, but not identical to, the cost function typically employed in geophysical parameter estimation techniques. The new method leads to the familiar Kalman Filter updates for the state, but also to an additional equation for estimating the parameter. This new equation is challenging to solve because it is nonlinear and involves derivatives of the forecast covariance matrix with respect to the parameter. Nevertheless, we were able to develop practical methods for solving these equations.

To illustrate the method, consider a stochastic version of the Lorenz three-variable model:

$$dx = -a(x - y)dt - \alpha_m(x - y) \circ dw \tag{2}$$

$$dy = (rxz - y)dt \tag{3}$$

$$dz = (xy - bz)dt \tag{4}$$

where dw is a Wiener process with zero mean and unit variance, and the "$\circ$" symbol indicates that the multiplicative noise is to be interpreted in the Stratonovich sense. We use the traditional parameter values for chaotic dynamics, namely $a = 10$, $b = 8/3$, and $r = 28$. In all experiments, the "truth" is defined as a single integration of the model starting from a randomly selected initial condition, and the "observation" is defined as the truth plus random numbers drawn independently from a normal distribution with zero mean and variance 0.01. We perform three distinct data assimilation experiments. First, we set α_m = 0 and assume that a is unknown and estimate its value using augmentation methods with the Ensemble Kalman Filter. The resulting estimates are shown in fig. 6a for three different initial conditions. The figures shows that the EnKF accurately estimates the parameter a independent of initial condition. In the second experiment, we set $a = 10$ and assume that α_m is unknown and estimate its value using the same augmentation method. The result, shown in fig. 6c, shows that this approach fails to converge to the correct value of the stochastic parameter; in fact, the converged value depends strongly on the initial condition (as can be anticipated theoretically). The difference in performance revealed in figs. 6a and 6c reflects the fact that a is a *deterministic* parameter and α_m is a *stochastic* parameter. Finally, in the third experiment, we again estimate α_m but this time using the newly developed Bayesian Filter. The result, shown in fig. 6d, shows that the filter gives reasonable estimates of the stochastic parameter. (Figure 6b shows a simplified version of the full Bayesian solution.) Also shown are the rank histograms of the probabilistic forecasts based on the augmented EnKF (fig. 6e) and Bayesian estimate (fig. 6f), which shows that the Bayesian methods produces "flatter" (and hence more reliable) probabilistic forecasts than the augmented EnKF. The results of this specific research project have been submitted for publication (DelSole and Yang, 2009).

5 Papers Generated by the Present Research Project

For convenience, we list here the papers that were generated by the present research project:

- Yang et al. (2008)
- Yang and DelSole (2009a)
- Yang and DelSole (2009b)
- DelSole and Yang (2009)

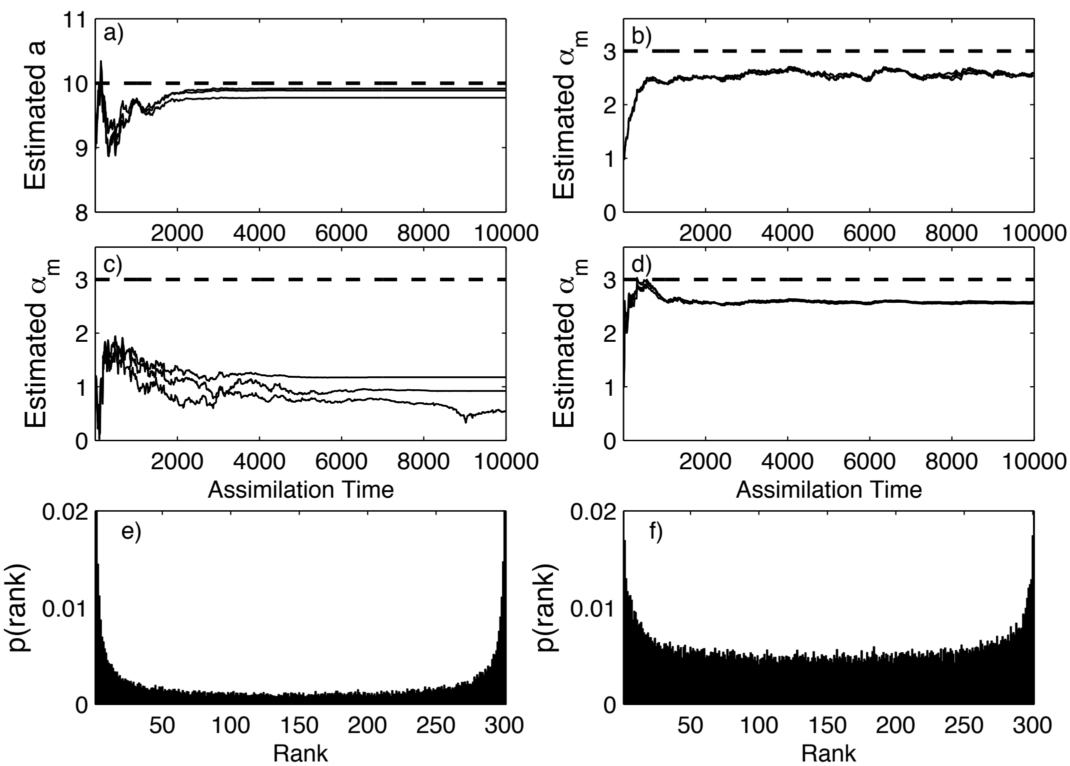

Figure 6: Results from four different data assimilation experiments with the Lorenz 3-variable model (2)-(4). Specifically, the figure shows as a function of assimilation time: (a) the estimated value of the parameter a using the augmentation method, (b) estimated value of α_m using the Bayesian method with no parameter variance update, (c) estimated value of α_m using the augmentation method, (d) estimated value of α_m using the Bayesian method with parameter variance update. In each case, three assimilation experiments were performed using the same initial condition but different realizations of noise. The true values are plotted as dashed lines for comparison. Also shown are (e) the rank histograms of x when α_m is estimated using the augmentation method (i.e., the experiment corresponding to panel (c)), and (f) the rank histograms of x when α_m is estimated using the Bayesian method with parameter variance update (i.e., the experiment corresponding to panel (d)).

References

Berner, J., F. J. Doblas-Reyes, T. N. Palmer, G. Shutts, and A. Weisheimer, 2008: Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model. *Phil. Trans. R. Soc. A*, **366**, 2561–2579.

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66**, 603–626.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.

Danforth, C. M., E. Kalnay, and T. Miyoshi, 2007: Estimating and correcting global weather model error. *Mon. Wea. Rev.*, **135**, 281–299.

DelSole, T. and A. Y. Hou, 1999: Empirical correction of a dynamical model. Part I: Fundamental issues. *Mon. Wea. Rev*, **127**, 2533–2545.

DelSole, T. and X. Yang, 2009: A Bayesian method for estimating stochastic parameters. *Physica D*, submitted.

Johansson, A. and S. Saha, 1989: Simulation of systematic error effects and their reduction in a simple model of the atmosphere. *Mon. Wea. Rev.*, **105**, 37–56.

Saha, S., 1992: Response of the NMC MRF model to systematic-error corection within integration. *Mon. Wea. Rev.*, **120**, 345–360.

Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102.

Smith, T. M. and R. W. Reynolds, 2004: Improved extended reconstruction of SST (1854-1997). *J. Climate*, **17**, 2466–2477.

Yang, X. and T. DelSole, 2009a: A diffuse ensemble square root filter. *Nonlin. Processes Geophys.*, **16**, 475–486.

Yang, X. and T. DelSole, 2009b: Using the ensemble Kalman Filter to estimate multiplicative model parameters. *Tellus*, **61**, 601–609.

Yang, X., T. DelSole, and H.-L. Pan, 2008: Empirical correction of the NCEP Global Forecast System. *Mon. Wea. Rev*, **136**, 5224–5233.

Yang, X.-Q. and J. L. Anderson, 2000: Correction of systematic errors in coupled gcm forecasts. *J. Climate*, **13**, 2072–2085.